

# **20th SIGBioMed Workshop on Biomedical Language Processing (BioNLP 2021)**

Online  
11 June 2021

ISBN: 978-1-7138-3017-7

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2021) by the Association for Computational Linguistics  
All rights reserved.

Printed with permission by Curran Associates, Inc. (2021)

For permission requests, please contact the Association for Computational Linguistics  
at the address below.

Association for Computational Linguistics  
209 N. Eighth Street  
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006  
Fax: 1-570-476-0860

[acl@aclweb.org](mailto:acl@aclweb.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>Improving BERT Model Using Contrastive Learning for Biomedical Relation Extraction</i> Peng Su, Yifan Peng and K. Vijay-Shanker .....	1
<i>Triplet-Trained Vector Space and Sieve-Based Search Improve Biomedical Concept Normalization</i> Dongfang Xu and Steven Bethard .....	11
<i>Scalable Few-Shot Learning of Robust Biomedical Name Representations</i> Pieter Fizez, Simon Suster and Walter Daelemans .....	23
<i>SAFFRON: tranSfer leArning For Food-disease RelatiOn extractiON</i> Gjorgjina Cenikj, Tome Eftimov and Barbara Koroušić Seljak .....	30
<i>Are we there yet? Exploring clinical domain knowledge of BERT models</i> Madhumita Sushil, Simon Suster and Walter Daelemans .....	41
<i>Towards BERT-based Automatic ICD Coding: Limitations and Opportunities</i> Damian Pascual, Sandro Luck and Roger Wattenhofer .....	54
<i>emrKBQA: A Clinical Knowledge-Base Question Answering Dataset</i> Preethi Raghavan, Jennifer J Liang, Diwakar Mahajan, Rachita Chandra and Peter Szolovits ...	64
<i>Overview of the MEDIQA 2021 Shared Task on Summarization in the Medical Domain</i> Asma Ben Abacha, Yassine Mrabet, Yuhao Zhang, Chaitanya Shivade, Curtis Langlotz and Dina Demner-Fushman .....	74
<i>WBI at MEDIQA 2021: Summarizing Consumer Health Questions with Generative Transformers</i> Mario Sängler, Leon Weber and Ulf Leser .....	86
<i>paht_nlp @ MEDIQA 2021: Multi-grained Query Focused Multi-Answer Summarization</i> Wei Zhu, Yilong He, Ling Chai, Yunxiao Fan, Yuan Ni, GUOTONG XIE and Xiaoling Wang ..	96
<i>BDKG at MEDIQA 2021: System Report for the Radiology Report Summarization Task</i> Songtai Dai, Quan Wang, Yajuan Lyu and Yong Zhu .....	103
<i>damo_nlp at MEDIQA 2021: Knowledge-based Preprocessing and Coverage-oriented Reranking for Medical Question Summarization</i> Yifan He, Mosha Chen and Songfang Huang .....	112
<i>Stress Test Evaluation of Biomedical Word Embeddings</i> Vladimir Araujo, Andrés Carvallo, Carlos Aspillaga, Camilo Thorne and Denis Parra .....	119
<i>BLAR: Biomedical Local Acronym Resolver</i> William Hogan, Yoshiki Vazquez Baeza, Yannis Katsis, Tyler Baldwin, Ho-Cheol Kim and Chun-Nan Hsu .....	126
<i>Claim Detection in Biomedical Twitter Posts</i> Amelie Wüthrl and Roman Klinger .....	131
<i>BioELECTRA: Pretrained Biomedical text Encoder using Discriminators</i> Kamal raj Kanakarajan, Bhuvana Kundumani and Malaikannan Sankarasubbu .....	143
<i>Word centrality constrained representation for keyphrase extraction</i> Zelalem Gero and Joyce Ho .....	155

<i>End-to-end Biomedical Entity Linking with Span-based Dictionary Matching</i> Shogo Ujiie, Hayate Iso, Shuntaro Yada, Shoko Wakamiya and Eiji ARAMAKI .....	162
<i>Word-Level Alignment of Paper Documents with their Electronic Full-Text Counterparts</i> Mark-Christoph Müller, Sucheta Ghosh, Ulrike Wittig and Maja Rey .....	168
<i>Improving Biomedical Pretrained Language Models with Knowledge</i> Zheng Yuan, Yijia Liu, Chuanqi Tan, Songfang Huang and Fei Huang .....	180
<i>EntityBERT: Entity-centric Masking Strategy for Model Pretraining for the Clinical Domain</i> Chen Lin, Timothy Miller, Dmitriy Dligach, Steven Bethard and Guergana Savova .....	191
<i>Contextual explanation rules for neural clinical classifiers</i> Madhumita Sushil, Simon Suster and Walter Daelemans .....	202
<i>Exploring Word Segmentation and Medical Concept Recognition for Chinese Medical Texts</i> Yang Liu, Yuanhe Tian, Tsung-Hui Chang, Song Wu, Xiang Wan and Yan Song .....	213
<i>BioM-Transformers: Building Large Biomedical Language Models with BERT, ALBERT and ELECTRA</i> Sultan Alrowili and Vijay Shanker .....	221
<i>Semi-Supervised Language Models for Identification of Personal Health Experiential from Twitter Data: A Case for Medication Effects</i> Minghao Zhu and Keyuan Jiang .....	228
<i>Context-aware query design combines knowledge and data for efficient reading and reasoning</i> Emilee Holtzapple, Brent Cochran and Natasa Miskov-Zivanov .....	238
<i>Measuring the relative importance of full text sections for information retrieval from scientific literature.</i> Lana Yeganova, Won Gyu KIM, Donald Comeau, W John Wilbur and Zhiyong Lu .....	247
<i>UCSD-Adobe at MEDIQA 2021: Transfer Learning and Answer Sentence Selection for Medical Summarization</i> Khalil Mrini, Franck Dernoncourt, Seunghyun Yoon, Trung Bui, Walter Chang, Emilias Farcas and Ndapa Nakashole .....	257
<i>ChicHealth @ MEDIQA 2021: Exploring the limits of pre-trained seq2seq models for medical summarization</i> Liwen Xu, Yan Zhang, Lei Hong, Yi Cai and Szui Sung .....	263
<i>NCUEE-NLP at MEDIQA 2021: Health Question Summarization Using PEGASUS Transformers</i> Lung-Hao Lee, Po-Han Chen, Yu-Xiang Zeng, Po-Lei Lee and Kuo-Kai Shyu .....	268
<i>SB_NITK at MEDIQA 2021: Leveraging Transfer Learning for Question Summarization in Medical Domain</i> Spandana Balumuri, Sony Bachina and Sowmya Kamath S .....	273
<i>Optum at MEDIQA 2021: Abstractive Summarization of Radiology Reports using simple BART Finetuning</i> Ravi Kondadadi, Sahil Manchanda, Jason Ngo and Ronan McCormack .....	280
<i>QIAI at MEDIQA 2021: Multimodal Radiology Report Summarization</i> Jean-Benoit Delbrouck, Cassie Zhang and Daniel Rubin .....	285

<i>NLM at MEDIQA 2021: Transfer Learning-based Approaches for Consumer Question and Multi-Answer Summarization</i>	
Shweta Yadav, Mourad Sarrouiti and Deepak Gupta .....	291
<i>IBMResearch at MEDIQA 2021: Toward Improving Factual Correctness of Radiology Report Abstractive Summarization</i>	
Diwakar Mahajan, Ching-Huei Tsou and Jennifer J Liang .....	302
<i>UETrice at MEDIQA 2021: A Prosper-thy-neighbour Extractive Multi-document Summarization Model</i>	
Duy-Cat Can, Quoc-An Nguyen, Quoc-Hung Duong, Minh-Quang Nguyen, Huy-Son Nguyen, Linh Nguyen Tran Ngoc, Quang-Thuy Ha and Mai-Vu Tran .....	311
<i>MNLP at MEDIQA 2021: Fine-Tuning PEGASUS for Consumer Health Question Summarization</i>	
Jooyeon Lee, Huong Dang, Ozlem Uzuner and Sam Henry .....	320
<i>UETfishes at MEDIQA 2021: Standing-on-the-Shoulders-of-Giants Model for Abstractive Multi-answer Summarization</i>	
Hoang-Quynh Le, Quoc-An Nguyen, Quoc-Hung Duong, Minh-Quang Nguyen, Huy-Son Nguyen, Tam Doan Thanh, Hai-Yen Thi Vuong and Trang M. Nguyen .....	328